Draft for Comments

Chapter: Synchronization

Text Book on

Fundamentals of Multimedia Computing

Authors:
Gerald Friedland
and
Ramesh Jain

Draft for Comments

Chapter: Synchronization

Introduction

A multimedia system uses data and information from multiple sensors to achieve its goals. Let us assume that $S_1, \ldots S_n$ are data streams from n different sensors of K types of data in the form of image sequence, audio stream, photos, motion detector, and other types including text. Each data stream has $M_1, \ldots, M_n$, as its metadata that may include location and type of the sensor, viewpoint, angles, camera calibration parameters or other similar parameters relevant to the data stream. As discussed earlier, a fundamental difference in multimedia from single medium understanding is that partial information from multiple media is correlated and combined to get complete information. Without correlating the information from multiple data streams, one can not extract information about the real world. Even in those systems, where multimedia is for direct consumption by humans, all correlated information must be presented for humans to extract information that they need from the multimedia data.

Synchronization is not a new problem. Whenever there is coordination among two or more sources is required to complete a task, coordination among those is required. In many cases this coordination becomes establishing, specifying, and then performing this coordination in space and time precisely. Two common examples of these are orchestra and any team sports. In soccer or football two or more players must be in a particular spatial configuration and must perform their roles perfectly for desirable outcome. In music, space is not that important but multiple players must produce particular sounds at specific instants for effect. In fact the term orchestrate means "*Arrange or direct the elements of (a situation) to produce a desired effect*". A music conductor uses specifications in a music book to coordinate the timing by individual players. As we will see in this chapter, synchronizations is to specify timing and location of each multimedia stream to create a holistic impact.

A common first step in extracting information from multiple streams as well as presenting information for final consumption from multiple streams is to make sure that they represent the same event although from different perspective and maybe using different modality. Since in most cases, the data from different sensors is collected independently and is usually transmitted and stored using different channels, care needs to be taken that all the data is synchronized. The process of synchronization usually referes to processes that are required to make sure that events in a multimedia system operate in the same unison as they do in the real world. The relationships that exist in the real world among different media components should be captured and maintained for information extraction as well as proper rendering for human experience. As we will see, the most important relationship that needs to be maintained is time, but spatial relationships and content relationships also need to be maintained in many applications.

In this chapter, we will discuss how these synchronizations affect and different techniques that are used to ascertain that such relationships are maintained.

In the following we first discuss content synchronization, then spatial synchronization. After these two concepts, we discuss temporal synchronization in more depth. It must be mentioned that temporal synchronization is the main synchronization approach. Much of the efforts spent in multimedia synchronization are on temporal synchronization because time plays very important role in rendering time dependent media. As we will see, increasingly people present even time-independent media such as text or photos in a video environment, where time plays a key role.

Content Synchronization

In many applications different types of data sets are semantically or functionally related to each other and make sense only if they appear together. Appearing together may be in terms of space or time or both. The important relationship to be maintained is the content relationship and should be rendered such that it makes sense. A common example could be appearance of a figure, or in some cases a text box or even a slide, close to the concept that it is related to. Embedding of figures close to their citation is commonly used. Increasingly, people are developing approaches to embed slide presentations along with a scientific paper. In all these techniques the emphasis is that one must present contents that are related to each other somehow close to each other.

It is very common that researchers present their work in a research article as well as a slide presentation that they may have made at a professional meeting. Usually, these two constitute a dual view of the same work, often quite different from each other. Slides represent help grasp the work at a high level and research paper presents detailed arguments. Since these two modes of presentations constitute a dual view, further utility can be gained if the two media are synchronized. In such a synchronized fine-grained alignment between slides and document passages could be constructed and presented, allowing a user to view both the slides and the document simultaneously.

This joint presentation of slides and a document can be prepared by finding a suitable fine-grained synchronization between them. Such a synchronization may be formulated as a problem of aligning slides to a corresponding paragraph span in the document. This problem was formalized in the Slideseer system as document-to- presentation alignment:

“Given a slide presentation S consisting of slides s1 to sn and a document D consisting of text paragraphs d1 to dm, an alignment is a function f(s) = (x, y), mapping each slide to a contiguous set of document paragraphs, starting at dx and ending at dy where x < = y, or to nil.”

The results of such an alignment are shown in Figure 1.

This is just an example of content synchronization. One could consider many situations, and the number of situations is increasing rapidly with increasingly availability and discovery methods on the Web. In most situations, the first step is to discover the content, then align it and finally present it.

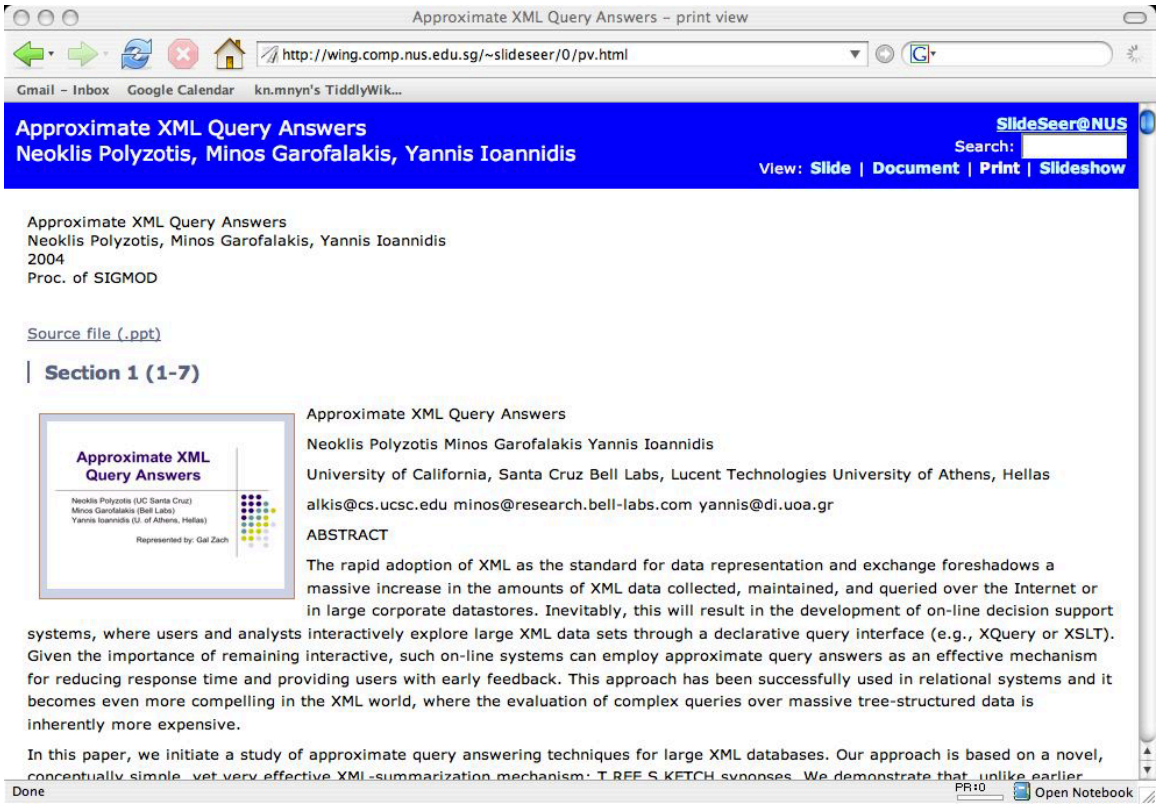

Figure 1: Alignment of slides and a research publication can be done analyzing contents of both and synchronizing them as reported in the Slideseer system [Reference].

Temporal Synchronization

Since many multimedia components are captured as time-varying signals and are also displayed in time, majority of the synchronization techniques have addressed issues in temporal synchronization. In this section, for brevity, we will drop 'temporal' unless needed in the context.

Synchronization techniques specify relationships that must be maintained among different media elements that must be interpreted or rendered together. In some cases all media elements are time-dependent, while in other cases such as making a video, some elements may not be time-dependent, but must be rendered at a specific time.A time-dependent object is a stream in which each element has a specified time associated with

it, while in a time-independent medium there may not be a stream (such as a photo or a text box) or the time relations may not be as critical, as in slide deck. Time-dependent object is presented or rendered as a media stream because there exists temporal relation between consecutive units of the stream. Time-independent object is the traditional medium such as text or images and could be rendered independent of time, except when they become part of a time-dependent media stream. Temporal dependencies among multiple media objects specify temporal relationships among objects that must be maintained for correct interpretation or understanding. A common example is the lip synchronization used in making movies. The image sequence stream showing a video must have a very precise temporal relationship with the audio steam representing speech for understanding what a person is saying. If the correct relationship is not maintained then a viewer may find the experience either poor or utterly confusing.

In many applications, temporal synchronization includes relations between time-dependent and time-independent objects as well. A slide show usually includes temporal synchronization of audio stream, either music or narration or both, which is time-dependent and individual slides which are time-independent media objects. In Figure 2, we show a time line and many different media objects that may be used to create a presentation or rendering of media as a function of time. For each media object its time duration is shown as a box on timeline. It is possible that at a given time multiple media objects may be rendered or at some times, no media object maybe rendered. The temporal relationships among time durations (commonly called time intervals) of different media objects may be specified using relationships first specified by Allen[XX]. Allen considered two processes and defined thirteen possible temporal relationships among them, see in Figure 3. These relationships have been used in specification of temporal relationship in many applications.

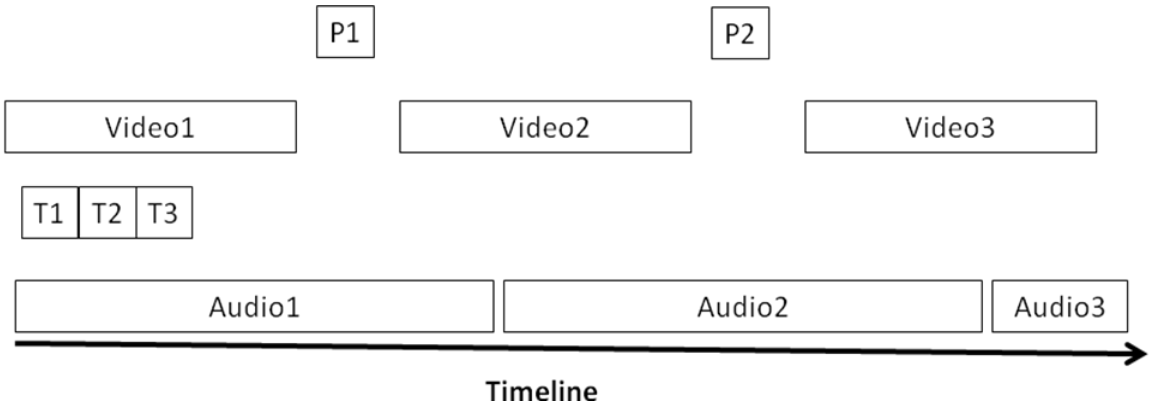

Figure 2: The duration of each media element, Audio, Video, text (T), and photo (P) are shown on the timeline.

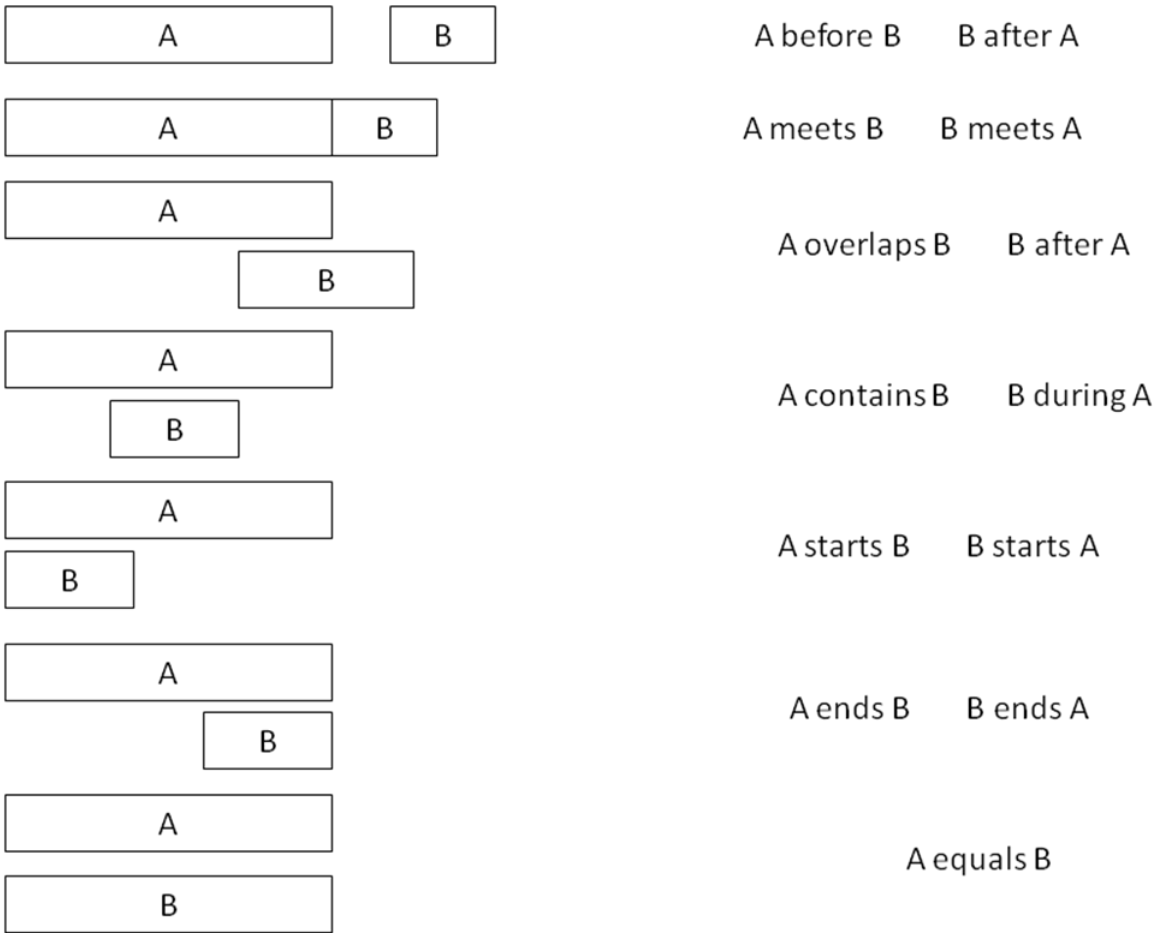

Figure 3: Thirteen temporal relations between two intervals A and B are shown here. Six of these are inverse of each other and hence there are only 7 relationships shown in this figure.

Synchronization Levels in a System

Synchronization has to be implemented at multiple levels in a system. One considers three common levels of synchronization.

Level 1: At the lowest level, support must be provided to display each stream smoothly. Operating systems and lower communication layers are responsible for ascertaining smooth single stream by bundling and making sure about the display requirements. They make sure that there is minimal jitter at the presentation time of the steam.

Level 2: The next level is commonly called the *RUN-TIME* support for synchronization of multimedia streams (schedulers) and is implemented on top of level 1. This level makes sure that skews between various streams is bounded to acceptable limits dictated by applications.

Level 3: The top level deals with the run-time support for synchronization between time-dependent and time-independent media. This level is also responsible for handling of user interaction. The main responsibility of this level is to make sure that the skews between time-dependent and time-independent media are within acceptable limits.

Specification of Synchronization

There are many ways to specify synchronization requirements among different media objects. In this section we discuss different specification approaches that may be used in applications depending on the type of media used and application requirements.

Implicit specification: In many applications, temporal relations among different media components are determined implicitly during their capture. The goal of a presentation of these objects is to present media in the same way as they were originally captured. The relationships among media objects is considered specified implicitly at the capture time. A common example of this is the capture of a video that consists of two media components: audio stream and image frame stream. These two streams are captured by the system using the same time line. Thus we may consider that the synchronization requirements are specified implicitly.

Explicit specification: Temporal relation may be specified explicitly in the case of presentations that are composed of independently captured time-dependent objects such as audio or video, time-independent objects such as photo, and manually created objects such as text-boxes and slides. Similarly, in applications like a slide show a presentation designer selects appropriate slides, selects audio objects, and defines relationships between the audio presentation stream and slides for presentation. In such cases, the relationships among different media objects must be explicitly defined and specified by the designer.

Intra-object Specification: An animation video comprises of all manually created image frames that are presented at appropriate frame rate to create a video. In all such cases, thought there is only one media stream, the objects in each frame in the sequence must be drawn to create specific visual effects. This requires that considering the motion characteristics to be conveyed, relationships and positions of different objects in each frame must be specified precisely. Thus, one needs to consider synchronization of the objects, their attributes, and their positions in each frame as a function of time in the presentation stream.

Deadlines

For synchronizing two streams, applications may specify whether their requirements for starting or ending media are rigid or flexible. In case of rigid requirements, the system

must make sure that the deadlines are considered hard and must meet specification. Soft or flexible requirements mean that there is some tolerance specified within which the relationship must be maintained. For example, in case of speech related audio and corresponding video showing a person's face, it is important that audio and video are played within a specified interval.

Spatial Synchronization

Spatial placement of different components of a document have played important role. As we saw in multimedia authoring, relative placement of a photo and its caption is important and must be clearly specified. If a caption and its corresponding figure do not appear together then it may result in a significant confusion. In fact, in many cases the semantics of the document or a simple figure may completely depend on the relative placement of its component. In Figure 4a, we show many components in a random order. By placing them appropriately with respect to each other in a spatial configuration, we get Figure 4b, which represents a stick figure representing a person. Without spatial relationships among different components, we will not be able to render the message that we want to communicate. In fact, one can easily see that by a different spatial arrangement of components, one will get a different object.

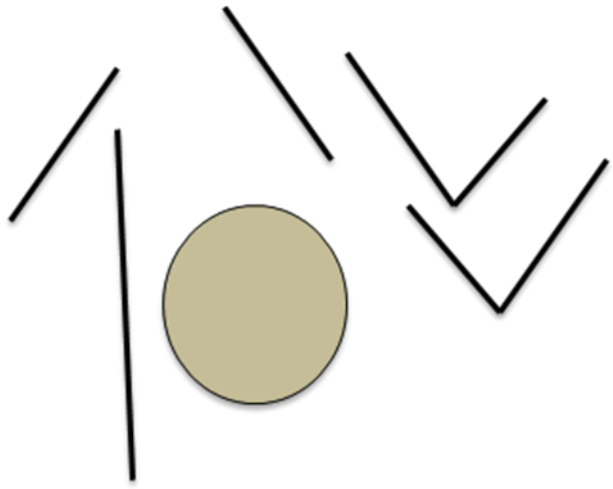

Figure 4a: Several components displayed in a random order.

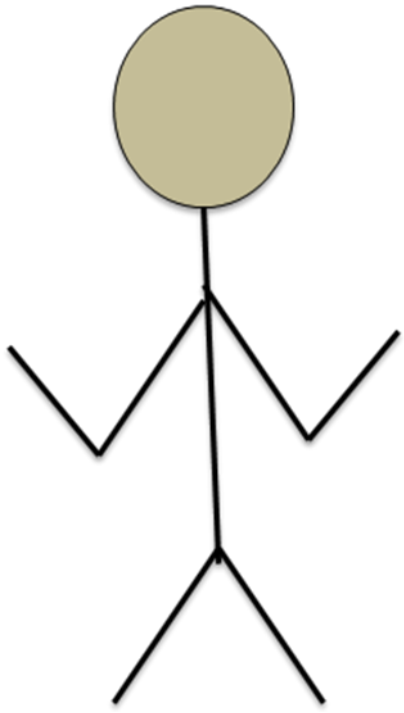

Figure 4b: The same components displayed in a specific order.

Rendering of different components of media is usually on a screen that is two dimensional. In some cases, one may want to consider rendering to take place in three dimensions, but in this chapter we will limit our discussion to two dimensional screens. On this screen, one must specify location of each media component as well as locations of sub components in each specific media.

A first step towards specifications is to define space used for presentation of a media object on an output device at a certain point of a time in a MM presentation. For this, ***Layout frames*** are defined on an output device, usually a screen, and each content is assigned to be inside this frame. Positioning of the layout frame is usually the size of the screen used by the application. Within this frame, windows corresponding to different media (photos, video, or even audio symbols) to be displayed could be defined. In many systems, the layout frame usually corresponds to the complete screen and the positions of the other windows are then defined relative to the layout frame.

In defining relative layouts and even objects within each media windows, it is useful to define more than one coordinate system. For example in Figure 5, we have defined three coordinate systems. We have a coordinate system, F, defined for the layout frame. Each media window is defined inside this window so the objects within each window could be defined with respect to the coordinate system, W, defined for it. Finally, components of an object in a window could be defined in coordinate system, O, defind for a specific

object in the window. By using multiple coordinate systems, and defining relationships among them, it is easier to define relationships that must be maintained with respect to different components.

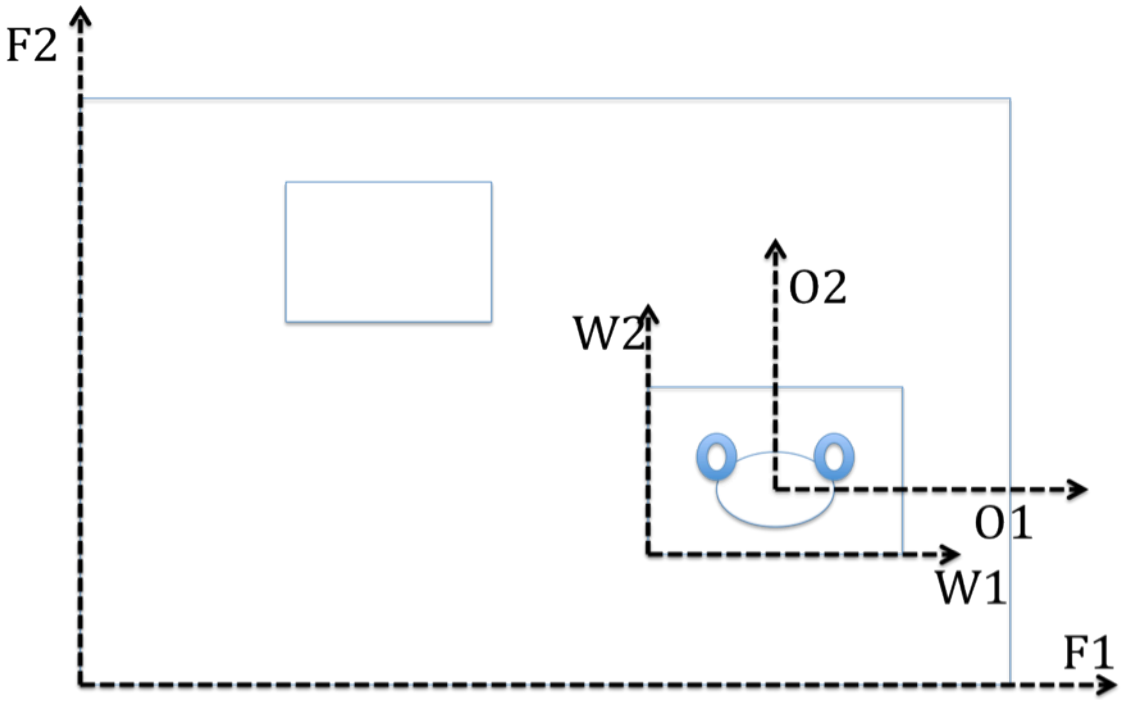

Figure 5: Use of multiple coordinate systems to specify positions among objects that must be maintained among them. The dashed lines represent three different coordinate systems F, W, and O, for Frame, Window, and Object, respectively.

REFERENCES

Allen, J.F., “Maintaining Knowledge about Temporal Intervals,” Comm. of the ACM, November 1983, Vol. 26, No. 11, pp. 832-843.

Min-Yen Kan (2007) SlideSeer: A Digital Library of Aligned Document and Presentation Pairs, In Proceedings of the Joint Conference on Digital Libraries (JCDL '07). Vancouver, Canada, June.

Reference on multiple coordinate systems.